

A Note on Polarization Martingales

Erdal Arıkan

Bilkent University

Ankara, Turkey

Email: arikan@ee.bilkent.edu.tr

Abstract—Polarization results rely on martingales of mutual information and entropy functions. In this note an alternative formulation is considered where martingales are constructed on sample functions of the entropy function.

I. INTRODUCTION

Let $(X, Y) \sim p_{X,Y}(x, y)$ denote a pair of discrete random variables with $x \in \mathcal{X} = \{0, 1\}$ and $y \in \mathcal{Y}$. In a channel coding context, one may think of X as the input to a binary-input channel and $p_{Y|X}$ as the channel transition probabilities. Alternatively, in the context of source coding, one may think of X as the output of a binary source and Y as side information about X .

Channel and source polarization results presented in [1], [2] were based the observation that if one takes two independent samples (X_1, Y_1) and (X_2, Y_2) from (X, Y) and defines $(U_1, U_2) = (X_1 + X_2, X_2)$, then one has

$$H(U_1|Y_1, Y_2) + H(U_2|Y_1, Y_2, U_1) = 2H(X|Y) \quad (1)$$

and

$$Z(U_1|Y_1, Y_2) + Z(U_2|Y_1, Y_2, U_1) \leq 2Z(X|Y). \quad (2)$$

The first relation (1) is simply the chain rule. The second relation (2) is proved in [2]. The above relations are the basis of source and channel polarization results. In order to obtain polarization, one simply extends the transform $(U_1, U_2) = (X_1 + X_2, X_2)$ recursively to higher orders and considers the entropy and Bhattacharyya terms generated in the course of the extension. In the extended transform, the relation (1) gives rise to a martingale in terms of the entropy function and (2) gives rise to a supermartingale defined in terms of the Bhattacharyya distance.

This note formulates polarization using a martingale on a per-sample basis. For $x \in \mathcal{X}$, $y \in \mathcal{Y}$, define the conditional entropy

$$h(x|y) = \log_2 \frac{1}{p_{X|Y}(x|y)}$$

This function represents samples of the conditional entropy random variable $h(X|Y)$, and the conditional entropy is given by $H(X|Y) = E[h(X|Y)]$. The samples of the conditional entropy random variable satisfy the chain rule (1) in the sense that

$$h(u_1|y_1, y_2) + h(u_2|y_1, y_2, u_1) = 2h(x|y). \quad (3)$$

This suggests that polarization martingales may be set up on a per sample basis. This note is a preliminary study in this direction.

Notation. Throughout, G_N denotes the polarization transform, which is defined for any $N = 2^n$, $n \geq 1$, by setting $G_N = F^{\otimes n} B_N$ where $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, $F^{\otimes n}$ is the n th Kronecker power of F , and B_N is the “bit-reversal” permutation [1]. The notation a^N denotes a vector (a_1, \dots, a_N) of length N .

II. POLARIZATION

This section gives a summary of the basic polarization result and establishes the notation and framework for the rest of the discussion. The presentation follows the source polarization setting of [2].

Proposition 1 (Polarization). [2] *Let $(X, Y) \sim p_{X,Y}(x, y)$ be a pair of random variables, with X taking values in $\{0, 1\}$. For any $N = 2^n$, $n \geq 1$, let $U^N = X^N G_N$. Then, for any $\delta \in (0, 1)$, as $N \rightarrow \infty$,*

$$\frac{|\{i \in [1, N]: H(U_i|Y^N, U^{i-1}) \in (1 - \delta, 1]\}|}{N} \rightarrow H(X|Y)$$

and

$$\frac{|\{i \in [1, N]: H(U_i|Y^N, U^{i-1}) \in [0, \delta)\}|}{N} \rightarrow 1 - H(X|Y).$$

For a full proof of this theorem, we refer to [2]. Here, we will sketch the main ideas to set the stage for presenting the new results. To begin, one may observe that there is in effect a global entropy conservation law

$$\sum_{i=1}^N H(U_i|Y^N, U^{i-1}) = NH(X|Y). \quad (4)$$

This follows by using the chain rule to write $H(U^N|Y^N) = \sum_{i=1}^N H(U_i|Y^N, U^{i-1})$. Then, since G_N is invertible, one writes $H(U^N|Y^N) = H(X^N|Y^N)$, and finally, by the memoryless channel assumption, one has $H(X^N|Y^N) = NH(X|Y)$.

The identity (4) ensures that the entropy variables $H(U_i|Y^N, U^{i-1})$ that appear prominently in Prop. 1 obey a global conservation law that is consistent with their claimed asymptotic behavior. However, this global conservation alone is not sufficient to force the individual entropy terms to polarize to 0 or 1. To prove polarization, a more refined analysis of the evolution of the entropy terms is needed. This is accomplished by embedding the entropy terms into a process that evolves as a martingale.

To define the polarization martingale, consider an infinite binary tree with a root node at level 0 and 2^n nodes at level $n \geq 1$. Let each node in the tree be labeled with a pair (n, i)

where $n \geq 0$ indicates the level of the node in the tree and $i = 1, \dots, 2^n$ the position of the node at level n . Each node (n, i) in the tree is connected to two nodes at level $n + 1$, referred to as the children of node (n, i) . Let the labeling be such that the children of a node (n, i) are the nodes $(n + 1, 2i - 1)$ and $(n + 1, 2i)$. Now, define a random walk $\{B_n; n \geq 0\}$ in this tree that starts at the root and moves one level deeper at each integer time by flipping a fair coin. More precisely, the random walk is such that (i) $B_0 = (0, 1)$ and (ii) given that $B_n = (n, i)$, B_{n+1} equals $(n + 1, 2i - 1)$ or $(n + 1, 2i)$ with probability $1/2$ each.

Associate an entropy $H^{(n,i)}$ to each node (n, i) in the above tree by setting $H^{(0,1)} = H(X|Y)$ and $H^{(n,i)} = H(U_i|Y^{2^n}, U^{i-1})$ for $n \geq 1, i = 1, \dots, 2^n$. Define the entropy process $\{H_n; n \geq 0\}$ so that

$$H_n = H^{B_n}. \quad (5)$$

Thus, $\{H_n\}$ is defined as a random process that reads the entropy label of the node visited by the random walk $\{B_n\}$. In other words, when $B_n = (n, i)$, the entropy process takes the value $H_n = H^{(n,i)}$. The process $\{H_n; n \geq 0\}$ is a martingale, i.e.

$$E[H_{n+1}|B_0, \dots, B_n] = H_n, \quad (6)$$

which follows from the chain rule

$$H^{(n+1,2i-1)} + H^{(n+1,2i)} = 2H^{(n,i)}. \quad (7)$$

For $n = 0$, (7) is equivalent to (1). The polarization construction recursively propagates the chain rule to higher levels. A crucial aspect of the polarization process is that

$$H^{(n+1,2i-1)} \leq H^{(n,i)} \leq H^{(n+1,2i)}. \quad (8)$$

One may think of the entropy process $\{H_n\}$ as a game in which H_n represents the fortune of a player at stage n of the game. The game starts with the player having an initial fortune $H_0 = H^{B_0} = H(X|Y)$. At each stage of the game, the player flips a fair coin and moves to a new fortune level. The game is fair in the sense that (7) holds. Excluding the case for equality in (8), the game is non-trivial in the sense that as the game is played the player's fortune fluctuates. The player's fortune is bounded in the sense that $0 \leq H_n \leq 1$. Prop. 1 can be interpreted as claiming that with probability one the player's fortune is destined to approach 0 or 1 asymptotically. Such claims are proved in [1] and [2] by invoking general convergence results about bounded martingales.

In both [1] and [2], the analysis of the martingale $\{H_n\}$ is accompanied by an auxiliary process $\{Z_n; n \geq 0\}$ based on the Bhattacharyya parameters. This process is defined alongside the entropy process $\{H_n\}$ using the same tree on which $\{H_n\}$ is defined. In addition to the already defined entropy label $H^{(n,i)}$, each node (n, i) in the tree gets a second label $Z^{(n,i)}$ so that $Z^{(0,1)} = Z(X|Y)$ and $Z^{(n,i)} = Z(U_i|Y^{2^n}, U^{i-1})$ for $n \geq 1, i = 1, \dots, 2^n$. The $\{Z_n\}$ process is defined by setting

$$Z_n = Z^{B_n}. \quad (9)$$

It can be shown that the Bhattacharyya process is a supermartingale,

$$E[Z_{n+1}|B_0, \dots, B_n] \leq Z_n. \quad (10)$$

The supermartingale property follows from the inequality

$$Z^{(n+1,2i-1)} + Z^{(n+1,2i)} \leq 2Z^{(n,i)}, \quad (11)$$

which is equivalent to (2) for $n = 0$ and holds in general by the special recursive structure of the polarization construction. The supermartingale $\{Z_n\}$ plays an auxiliary role in helping determine the rate of convergence of the entropy process $\{H_n\}$ [1], [3].

III. PER-SAMPLE ENTROPY MARTINGALE

Source polarization in the previous section has been given in terms of a martingale $\{H_n; n \geq 0\}$ which took values in the space of entropies. Each such value $H(U_i|Y^N, U^{i-1})$ is the expectation of an entropy random variable $h(U_i|Y^N, U^{i-1})$. We now consider a more refined martingale formulation for polarization in which the martingale takes values in the space of sample functions of the entropy random variables, i.e., values of the form $h(u_i|y^N, u^{i-1})$. To distinguish the two types of martingales, we will refer to the new martingale as a *per-sample* martingale. The per-sample entropy martingale will be denoted as $\{h_t; 0 \leq t \leq n\}$. Unlike $\{H_n; n \geq 0\}$ that can be extended indefinitely in duration, the per-sample process will have a finite duration $n = \log_2(N)$ where N is the size of the sample. The sample size N can be an arbitrarily large power of two; but for a given N , the length of the per-sample process will be $\log_2(N)$.

To define the per-sample entropy process fix a sample size $N = 2^n$. Let (x^N, y^N) be a fixed but arbitrary sample. Let $u^N = x^N G_N$ and consider the entropy terms $h(u_i|y^N, u^{i-1})$, $i = 1, \dots, N$. Index these entropy terms using n bits. Specifically, denote the i th entropy term $h(u_i|y^N, u^{i-1})$ alternatively as $h^{(b_n, \dots, b_1)}$ where $(b_n, \dots, b_1) \in \{0, 1\}^n$ is the binary representation of the integer $(i - 1)$; in other words, $i - 1 = \sum_{j=1}^n b_j 2^{j-1}$. Let (B_1, B_2, \dots, B_n) be uniformly distributed over the index space $\{0, 1\}^n$. Define the terminal element of the per-sample entropy process $\{h_t; 0 \leq t \leq n\}$ by setting

$$h_n = h^{(B_n, \dots, B_1)}. \quad (12)$$

With this definition, h_n is equally likely to take on any of the sample values $h(u_i|y^N, u^{i-1})$, $i = 1, \dots, N$ (supposing all values are distinct). The remainder of the entropy process is defined as

$$h_t = \begin{cases} E[h_n], & t = 0; \\ E[h_n|B_1, B_2, \dots, B_t], & 1 \leq t \leq n. \end{cases} \quad (13)$$

The construction of $\{h_t; 0 \leq t \leq n\}$ follows Doob's method of generating a martingale from a given random variable (h_n in this instance) [5, p. 297]; hence, by construction, we have a martingale in the sense that

$$E[h_{t+1}|B_1, B_2, \dots, B_t] = h_t, \quad t = 0, 1, \dots, n-1. \quad (14)$$

It should be noted that the above construction creates a martingale regardless of the structure of the transform G_N . However, in the case of the specific polarization transform G_N that we have here, the resulting martingale has a recursive structure that can be represented by a simple circuit diagram. To gain further insight into the structure of the per-sample entropy process $\{h_t\}$ under the polarization transform, we will study a small example with the aid of Fig. 1. The figure shows a circuit for the polarization transform $u^N = x^N G_N$ with sample size $N = 8$. The labels on the rightmost side of the circuit are fixed by the sample (x^8, y^8) . Given x^8 , the circuit calculates the remaining variables $\{v_{t,i} : 1 \leq t \leq 3, 1 \leq i \leq 8\}$ in accordance with the usual computation rules implied by the wiring diagram. The transform u^N corresponding to x^N is obtained at the left-most side of the circuit as $u_i = v_{3,i}$. For notational uniformity, it will be convenient to define $v_{0,i} = x_i$, as well.

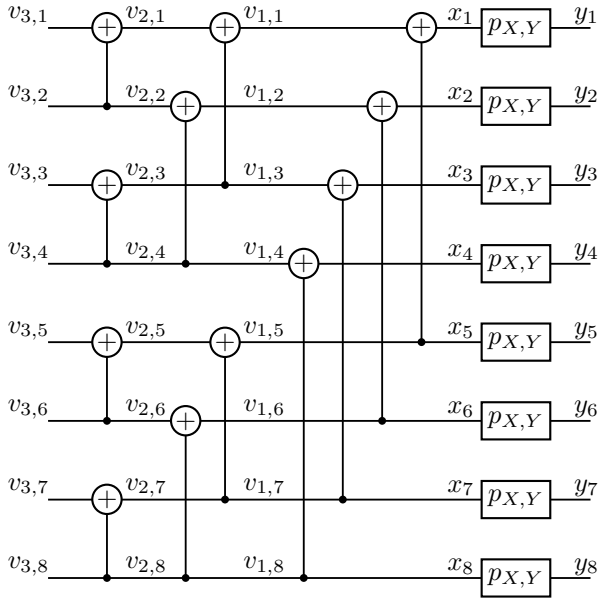


Fig. 1. Polarization transform of size $N = 8$.

The process $\{h_t; t = 0, 1, 2, 3\}$ starts at time $t = 0$ with the value $h_0 = E[h_3] = \frac{1}{8} \sum_{i=1}^8 h(u_i|y^8, u^{i-1}) = \frac{1}{8} h(u^8|y^8) = \frac{1}{8} h(x^8|y^8)$ where the last step follows from the fact that $u^8 = x^8 G_8$ is a 1-1 transform. The value h_0 is the mean conditional entropy for the given sample (x^8, y^8) .

At time $t = 1$, the value of the process is determined by the first index bit B_1 .

$$h_1 = \begin{cases} \frac{1}{4} h(v_{1,1}, \dots, v_{1,4} | y^8), & \text{if } B_1 = 0, \\ \frac{1}{4} h(v_{1,5}, \dots, v_{1,8} | y^8, v_{1,1}, \dots, v_{1,4}), & \text{if } B_1 = 1. \end{cases}$$

At time $t = 2$, the value is a function of the first two index

bits $B^2 = (B_1, B_2)$.

$$h_2 = \begin{cases} \frac{1}{2} h(v_{2,1}, v_{2,2} | y^8), & \text{if } B^2 = (0, 0), \\ \frac{1}{2} h(v_{2,3}, v_{2,4} | y^8, v_{2,1}, v_{2,2}), & \text{if } B^2 = (0, 1), \\ \frac{1}{2} h(v_{2,5}, v_{2,6} | y^8, v_{2,1}, \dots, v_{2,4}), & \text{if } B^2 = (1, 0), \\ \frac{1}{2} h(v_{2,7}, v_{2,8} | y^8, v_{2,1}, \dots, v_{2,6}), & \text{if } B^2 = (1, 1). \end{cases}$$

Finally, at time $t = 3$, the value of the process as a function of $B^3 = (B_1, B_2, B_3)$ is as follows.

$$h_3 = \begin{cases} h(v_{3,1} | y^8), & \text{if } B^3 = (0, 0, 0), \\ h(v_{3,2} | y^8, v_{3,1}), & \text{if } B^3 = (0, 0, 1), \\ h(v_{3,3} | y^8, v_{3,1}, \dots, v_{3,2}), & \text{if } B^3 = (0, 1, 0), \\ h(v_{3,4} | y^8, v_{3,1}, \dots, v_{3,3}), & \text{if } B^3 = (0, 1, 1), \\ h(v_{3,5} | y^8, v_{3,1}, \dots, v_{3,4}), & \text{if } B^3 = (1, 0, 0), \\ h(v_{3,6} | y^8, v_{3,1}, \dots, v_{3,5}), & \text{if } B^3 = (1, 0, 1), \\ h(v_{3,7} | y^8, v_{3,1}, \dots, v_{3,6}), & \text{if } B^3 = (1, 1, 0), \\ h(v_{3,8} | y^8, v_{3,1}, \dots, v_{3,7}), & \text{if } B^3 = (1, 1, 1). \end{cases}$$

Upon reaching level $n = 3$, the process stops at one of the specific sample values $h(u_i | y^N, u^{i-1})$.

The above observations can be summarized as follows.

Proposition 2. For any sample (x^N, y^N) of the ensemble $(X^N, Y^N) \sim \prod_{i=1}^N p_{X,Y}(x_i, y_i)$, the per-sample entropy process $\{h_t; 0 \leq t \leq n\}$ defined by (12) and (13) is a martingale in the sense of (14). The process remains a martingale if the sample value (x^N, y^N) is replaced by the random pair (X^N, Y^N) . For any fixed n , the entropy process $\{H_t; 0 \leq t \leq n\}$ defined by (5) is the mean of the per-sample entropy process $\{h_t; 0 \leq t \leq n\}$ in the sense that $H_t = E[h_t]$ where the expectation is w.r.t. the ensemble (X^N, Y^N) .

The convergence properties of the per-sample entropy martingale by large deviation techniques is a subject left for future study. Such an analysis is likely to be useful in the performance analysis of polar codes.

ACKNOWLEDGMENT

This work was supported in part by TÜBİTAK under grant 110E243 and in part by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMunications NEWCOM# (contract n.318306).

REFERENCES

- [1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 3051–3073, July 2009.
- [2] E. Arıkan, "Source polarization," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, pp.899,903, 13–18 June 2010.
- [3] E. Arıkan and E. Telatar, "On the rate of channel polarization" in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, Seoul, Korea, pp. 1493–1495, June 28–July 3, 2009.
- [4] S. B. Korada, *Polar codes for channel and source coding*. PhD thesis, EPFL, Lausanne, 2009.
- [5] S. M. Ross, *Stochastic Processes*, 2nd Ed. Wiley, 1996.